Amendments to the Claims

Please cancel Claims 4 and 13. Please amend Claims 1, 5, 10, and 14. The Claim Listing below will replace all prior versions of the claims in the application:

Claim Listing

1. (Currently amended) A computer-implemented method of determining content type of contents of a subject Web page, comprising the steps of:

providing a predefined set of potential content types, content types being exclusive of indicating formal language of the content;

for each potential content type, preparing a distinguishing series of tests, the distinguishing series of tests includes:

i) at least one binary test, and

ii) at least one non-binary binary test,

the at least one binary test and the at least one non-binary test further including at least one test (a) examining syntax or grammar; or (b) examining page format or style other than position of data or a keyword in the subject Web page;

for each potential content type, running the distinguishing series of tests having test results which enable quantitative evaluation of at least some contents of the subject Web page being of the potential content type;

mathematically combining the probabilities from all possible combinations of the test results and hypothesis values with respect to content of Web pages of determined content type with the test results of the subject Web page of undetermined content type using at least one Bayesian network; and

based on the combined test results, assigning a respective probability, for each potential content type, that some contents of that type exists on the subject Web page, and indicating content type, said indicating being exclusive of indicating language in which content is written.

2. (Original) A method as claimed in Claim 1 wherein the set of potential content types include any combination of organization description, organization history, organization

mission, organization products/services, organization members, organization contact information, management team information, job opportunities, press releases, calendar of events/activities, biographical data, articles/news with information about people, articles/news with information about organizations and employee roster.

3. (Original) A method as claimed in Claim 1 wherein the step of combining includes producing a respective confidence level for each potential content type, that at least some content of the subject Web page is of the potential content type.

4. (Cancelled)

5. (Currently amended) A method as claimed in Claim [[4]] 1 further comprising the step of training the Bayesian network using a training set of Web pages with respective known content types such that statistics on the test results are collected on the training set of Web pages.

6. (Previously presented) A method as claimed in Claim 1 wherein the predefined set includes a potential content type of press release and the distinguishing series of tests further includes at least one of:

(i) determining whether a predefined piece of data or keyword appears in the subject Web page;

(ii) examining text properties; or

(iii) determining whether the predefined piece of data or keyword appears in URLs in the subject Web page.

7. (Previously presented) The method as claimed in Claim 1 wherein the distinguishing series of tests further includes at least one of:

(i) determining whether a predefined piece of data or keyword appears in the subject Web page;

(ii) examining text properties; or

(iii) determining whether the predefined piece of data or keyword appears in URLs in the subject Web page.

8. (Original) A method as claimed in Claim 1 further comprising the step of storing indications of the assigned probabilities of each potential content type per respective Web page.

9. (Original) A database formed by the method of Claim 8, said database containing indications of Web pages and corresponding content types determined to be found on respective Web pages.

10. (Currently amended) Apparatus for determining content type of contents of a subject Web page, comprising:

a predefined set of potential content types, each potential content type being exclusive of indicating formal language of the content and associated with a respective distinguishing series of tests, the distinguishing series of tests includes:

i) at least one binary test, and

ii) at least one non-binary binary test,

the at least one binary test and the at least one non-binary test further including at least one test (a) examining syntax or grammar; or (b) examining page format or style other than position of data or a keyword in the subject Web page;

a test module utilizing the predefined set, the test module employing the distinguishing series of tests as a plurality of processor-executed tests having test results which enable, for each potential content type, quantitative evaluation of at least some contents of the subject Web page being of the potential content type, for each potential content type, the test module (i) running the respective distinguishing series of tests, (ii) combining the probabilities from all possible combinations of the test results and hypothesis values with respect to content of Web pages of determined content type with the test results of the subject Web page of undetermined content type using at least one Bayesian network and (iii) for each potential content type, assigning a respective

probability that at least some contents of that type exists on the subject Web page being of the potential content type, and indicating content type exclusive of indicating language in which content is written.

11. (Original) Apparatus as claimed in Claim 10 wherein the set of potential content types include any combination of contact information, press release, company description, employee list, other.

12. (Original) Apparatus as claimed in Claim 10 wherein the test module produces a respective confidence level for each potential content type, that at least some content of the subject Web page is of the potential content type.

13. (Cancelled)

14. (Currently amended) Apparatus as claimed in Claim [[13]] 10 further comprising a training member for training the Bayesian network using a training set of Web pages with respective known content types, such that statistics on the test results are collected on the training set of Web pages.

15. (Original) Apparatus as claimed in Claim 10 wherein the predefined set includes a potential content type of at least one of organization description, organization history, organization mission, organization products/services, organization members, organization contact information, management team information, job opportunities, press releases, calendar of events/activities, biographical data, articles/news with information about people, articles/news with information about organizations and employee roster.

16. (Previously presented) Apparatus as claimed in Claim 15 wherein the processor-executed tests include at least one of:

(i) determining whether a predefined piece of data or keyword appears in the subject Web page;

(ii) examining text properties; or

(iii) determining whether the predefined piece of data or keyword appears in URLs in the subject Web page.

17. (Previously presented) Apparatus as claimed in Claim 10 wherein the processor-executed tests include any of:

(i) determining whether a predefined piece of data or keyword appears in the subject Web page;

(ii) examining text properties; or

(iii) determining whether the predefined piece of data or keyword appears in URLs in the subject Web page.

18. (Original) Apparatus as claimed in Claim 10 further comprising storage means for receiving and storing indications of the assigned probabilities of each content type per Web page as determined by the test module, such that the storage means provides a cross reference between a Web page and respective content types of contents found on that Web page.

19. (Previously presented) A method as claimed in Claim 1 wherein the at least one binary test and the at least one non-binary tests include one or more of the following tests:

i) whether the subject Web page contains a press release;

ii) whether the subject Web page has a title;

iii) whether the subject Web page has a copyright statement;

iv) whether the subject Web page has a navigation map;

v) whether the subject Web page has a line with a keyword followed by at least another keyword within the next 10, 20, 30 or 40 lines;

vii) whether a first sentence of a first paragraph of the subject Web page has a date;

viii) whether the first sentence of the first paragraph of the subject Web page is preceded by a header line;

ix) whether the first sentence of the first paragraph of the subject Web page contains the keyword or a form of the keyword;

xi) whether the subject Web page contains a text line starting with the keyword; and

xii) a calculation of a percentage of header lines, the average sentence length, number of different domains, number of lines that contain the keyword or number of phrases that contain the keyword.

20. (Previously presented) Apparatus as claimed in Claim 10 wherein the at least one binary test and the at least one non-binary tests include one or more of the following tests:

i) whether the subject Web page contains a press release;

ii) whether the subject Web page has a title;

iii) whether the subject Web page has a copyright statement;

iv) whether the subject Web page has a navigation map;

v) whether the subject Web page has a line with a keyword followed by at least another keyword within the next 10, 20, 30 or 40 lines;

vii) whether a first sentence of a first paragraph of the subject Web page has a date;

viii) whether the first sentence of the first paragraph of the subject Web page is preceded by a header line;

ix) whether the first sentence of the first paragraph of the subject Web page contains the keyword or a form of the keyword;

xi) whether the subject Web page contains a text line starting with the keyword; and

xii) calculation of a percentage of header lines, the average sentence length, number of different domains, number of lines that contain the keyword or number of phrases that contain the keyword.

21. (Previously presented) The method of Claim 1, wherein the at least one test examining syntax or grammar includes at least one test measuring one of the following: number of

passive sentences, number of sentences without a verb, and percentage of verbs in past tense.

22. (Previously presented) The apparatus of Claim 10, wherein the at least one test examining syntax or grammar includes at least one test measuring one of the following: number of passive sentences, number of sentences without a verb, and percentage of verbs in past tense.